Data management protocols workshop Discipline specific protocol: rate your reproducibility

Kees den Heijer

Data Steward faculty Civil Engineering and Geosciences, TU Delft

University of Tartu
April 10, 2019

LaTeX source revision 173 (April 10, 2019)

Protocols

- What do you consider a protocol?
- Are protocols used in your field?

About us

Data Stewards TU Delft

Kees den Heijer

Data Steward
Faculty of Civil Engineering and Geosciences

Esther Plomp

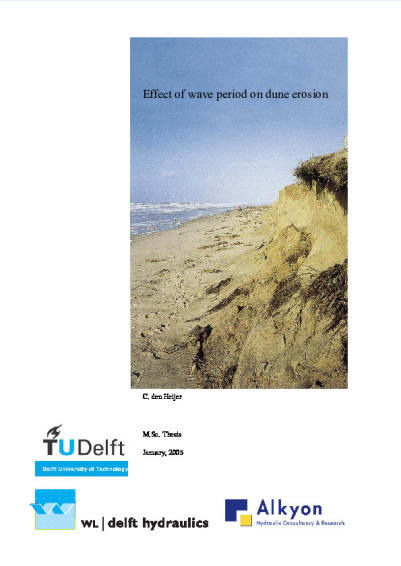
Data Steward
Faculty of Applied Sciences

About the speaker

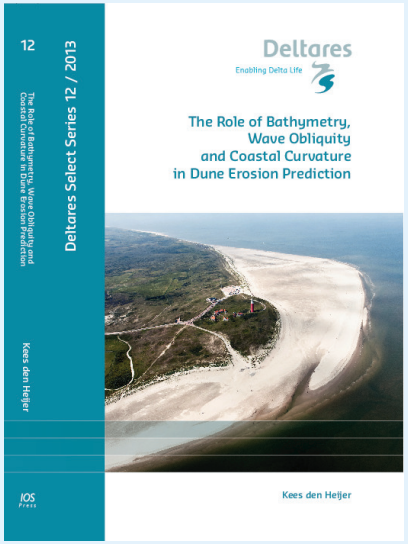
Experienced in data and information matter

- Data Steward @ CEG faculty
- Owner / consultant @ Data2day

Educated in Coastal Engineering (TU Delft, CEG)



MSc (2005)



PhD (2013)

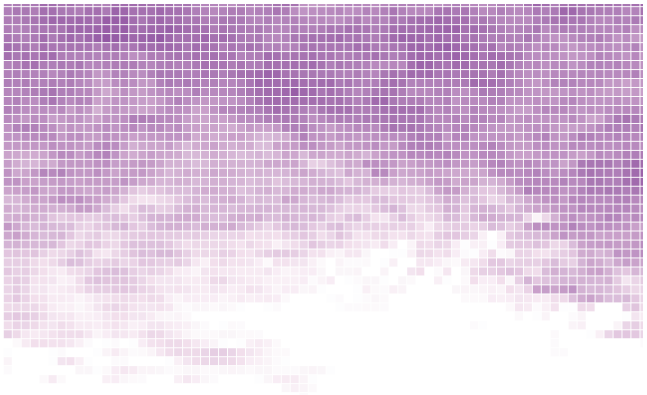
What is your data management like?

Outline

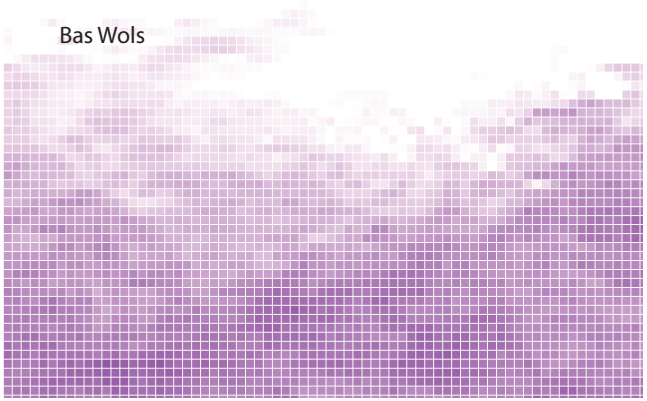
Discipline specific protocol: rate your reproducibility

Open research data

Computational Fluid Dynamics (CFD) in drinking water treatment (Wols, 2010)



CFD in drinking water treatment



Kees den Heijer

Open research data

Computational Fluid Dynamics (CFD) in drinking water treatment (Wols, 2010)

Data publications

Experimental and numerical data used in this thesis can be accessed online (through http://dx.doi.org).

All data:
`http://dx.doi.org/10.4121/uuid:c1ac7344-1419-4398-ba13-c757551c303f`

Movies of particle distributions in ozone systems (Figure 4.8):
`http://dx.doi.org/10.4121/uuid:f284a6f2-04a2-431e-9896-4a04bd61ac6b`

Movies of spreading of dye around a single cross-flow UV lamp (Figure 5.3 & 5.4):
`http://dx.doi.org/10.4121/uuid:954f0e25-aa39-454d-a0fa-656f36bf72d1`

Movies of advanced oxidation around a single cross-flow UV lamp (Figure 5.18):
`http://dx.doi.org/10.4121/uuid:dfff9d45-75c4-455c-848a-d50bc4a09087`

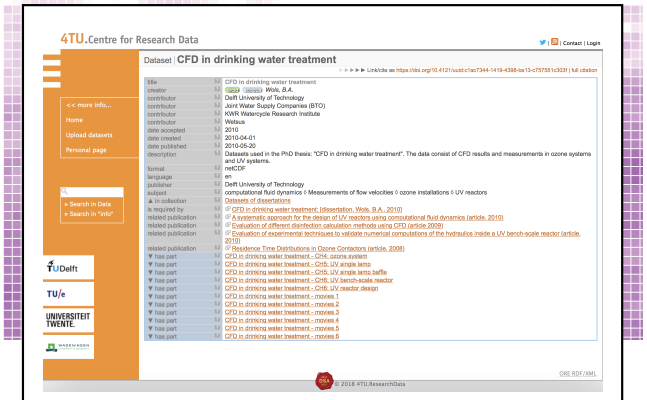
Movies of particle distributions in single cross-flow UV lamp systems (Figure 5.20):
`http://dx.doi.org/10.4121/uuid:63ef98c8-5a01-442a-9caf-80106576a171`

Movies of spreading of dye in a bench-scale UV reactor(Figure 6.4):
`http://dx.doi.org/10.4121/uuid:06d6b6e3-3fe1-4502-825b-6627a5deaa5f`

Movies of a LES model of a bench-scale UV reactor (Figure 7.1):
`http://dx.doi.org/10.4121/uuid:474d9ab9-c354-4d18-87c9-69b02acedce3`

CFD in drinking water treatment

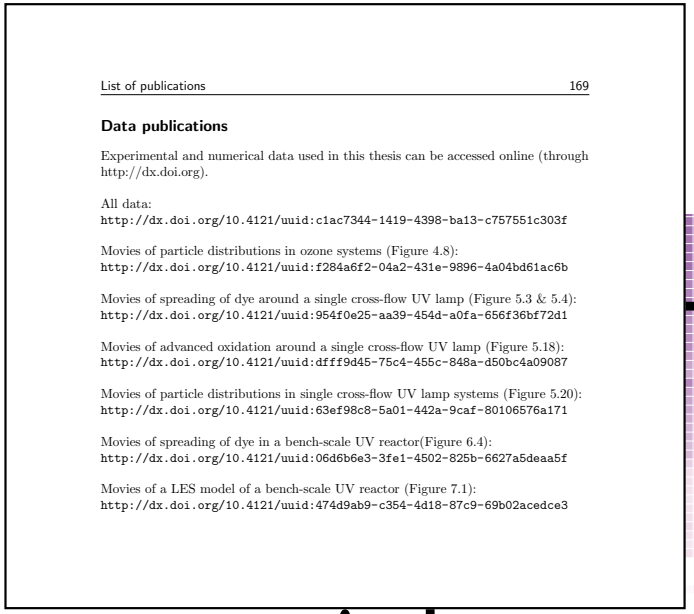
Bas Wols



Kees den Heijer

Open research data

Computational Fluid Dynamics (CFD) in drinking water treatment (Wols, 2010)



List of publications 169

Data publications

Experimental and numerical data used in this thesis can be accessed online (through http://dx.doi.org).

All data:
http://dx.doi.org/10.4121/uuid:c1ac7344-1419-4398-ba13-c757551c303f

Movies of particle distributions in ozone systems (Figure 4.8):
http://dx.doi.org/10.4121/uuid:f284a6f2-04a2-431e-9896-4a04bd61ac6b

Movies of spreading of dye around a single cross-flow UV lamp (Figure 5.3 & 5.4):
http://dx.doi.org/10.4121/uuid:954f0e25-aa39-454d-a0fa-656f36bf72d1

Movies of advanced oxidation around a single cross-flow UV lamp (Figure 5.18):
http://dx.doi.org/10.4121/uuid:dfff9d45-75c4-455c-848a-d50bc4a09087

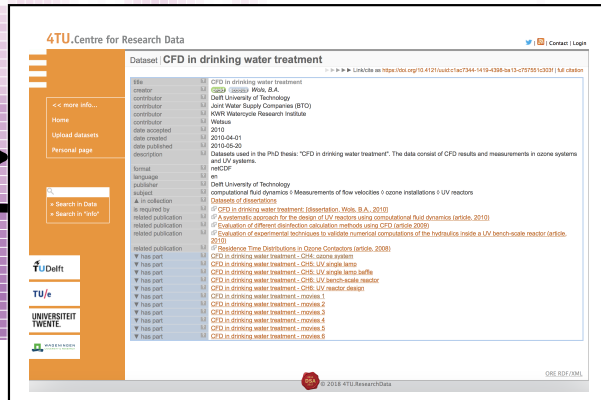
Movies of particle distributions in single cross-flow UV lamp systems (Figure 5.20):
http://dx.doi.org/10.4121/uuid:63ef98c8-5a01-442a-9caf-80106576a171

Movies of spreading of dye in a bench-scale UV reactor(Figure 6.4):
http://dx.doi.org/10.4121/uuid:06d6b6e3-3fe1-4502-825b-6627a5deaa5f

Movies of a LES model of a bench-scale UV reactor (Figure 7.1):
http://dx.doi.org/10.4121/uuid:474d9ab9-c354-4d18-87c9-69b02acedce3

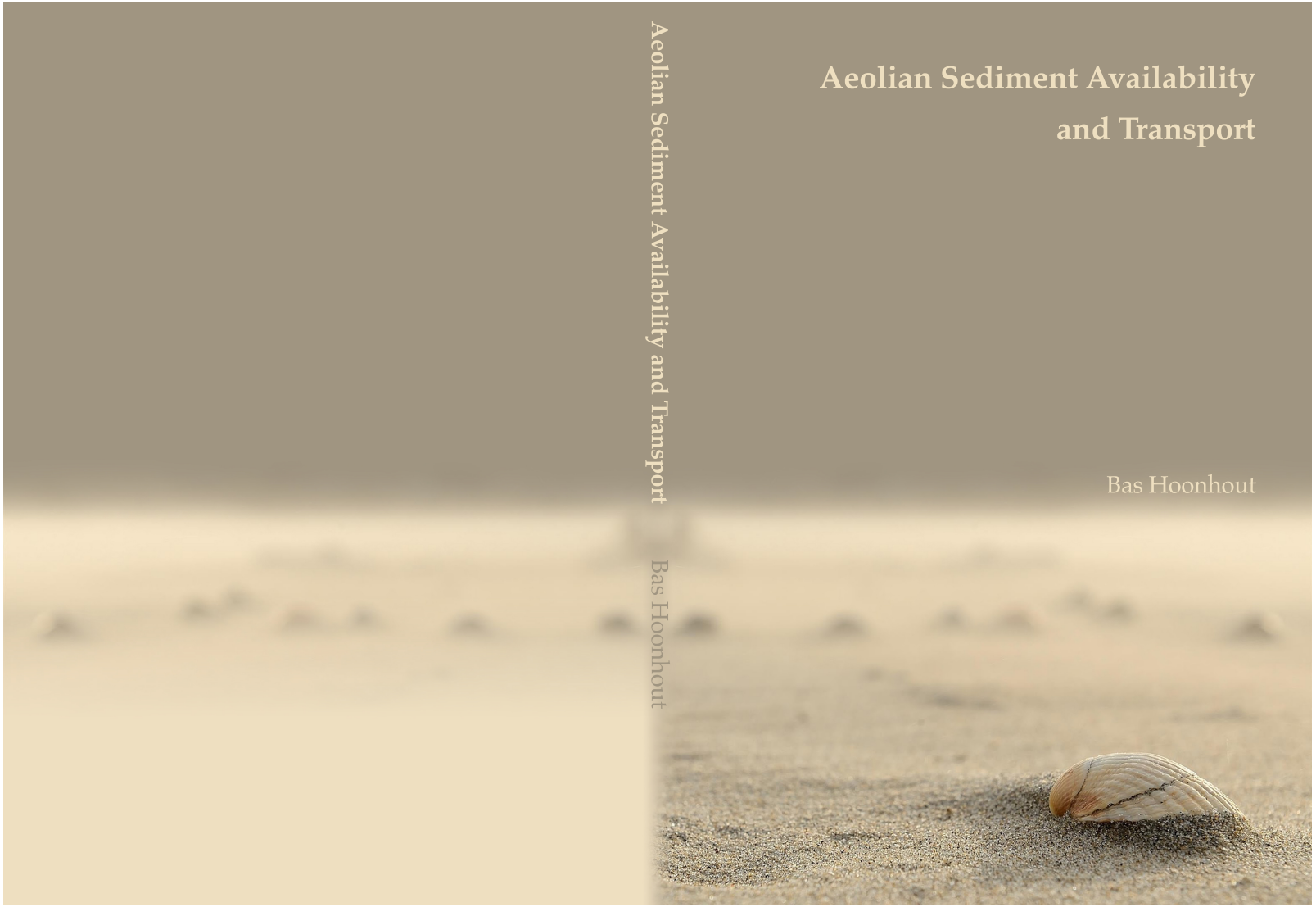
CFD in drinking water treatment

Bas Wols



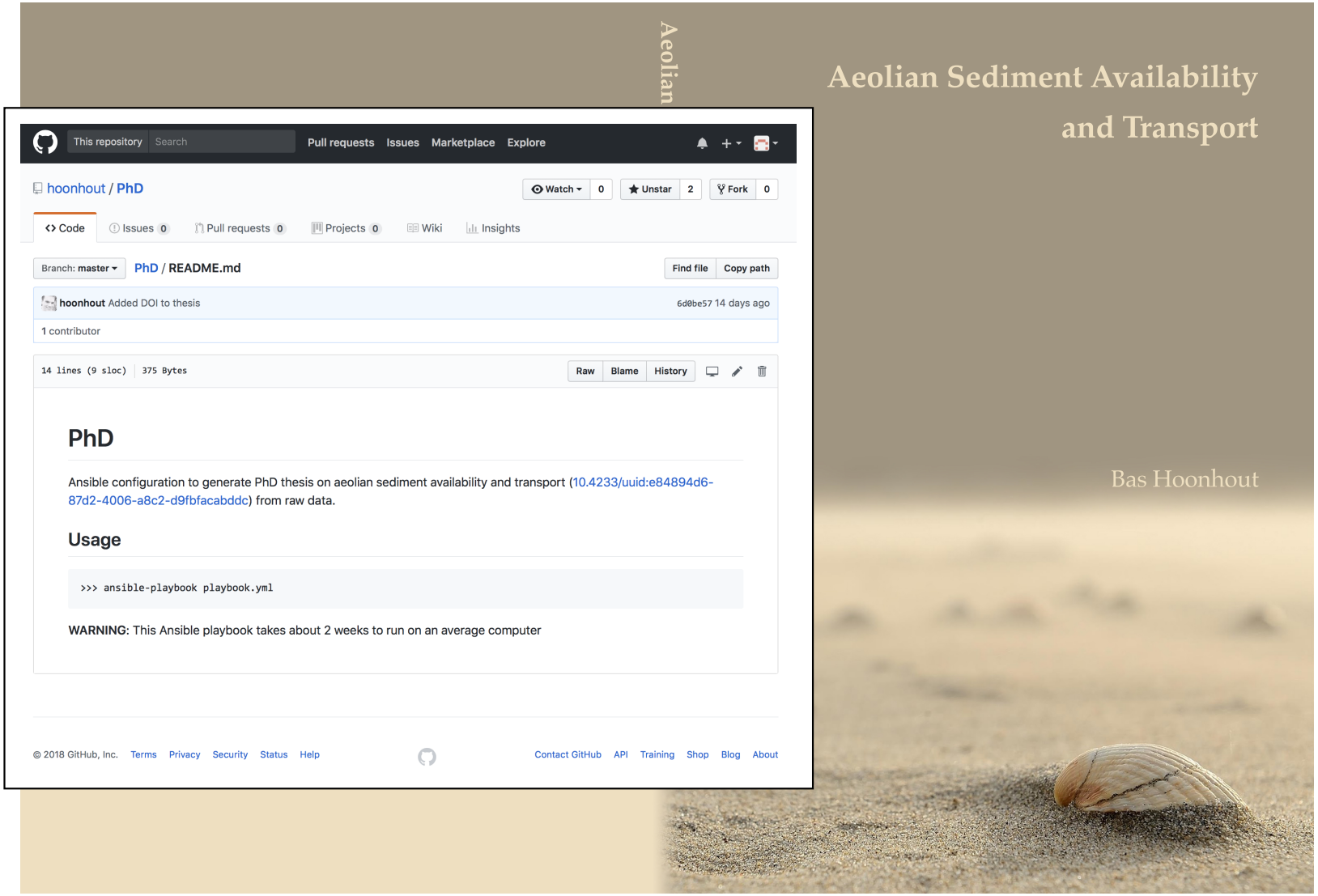
Open and reproducible research

Aeolian Sediment Availability and Transport (Hoonhout, 2017)



Open and reproducible research

Aeolian Sediment Availability and Transport (Hoonhout, 2017)



Sand Motor

Introduction

Large scale experiment

photo 2011, by Joop van Houdt

Construction year 2011

Sand volume 21.5 10^6 m^3

Research

- Various reseach grants
- 15 PhD researchers
- International cooperation
- Multi-disciplinary
  - Morphology
  - Biology
  - Ecology
  - Hydrology
  - Chemistry
  - Governance

Sand Motor

Introduction

Large scale experiment

photo 2017, by Joop van Houdt

Construction year 2011

Sand volume 21.5 10^6 m^3

Research

- Various reseach grants
- 15 PhD researchers
- International cooperation
- Multi-disciplinary
  - Morphology
  - Biology
  - Ecology
  - Hydrology
  - Chemistry
  - Governance

Types of data

Wide variety

OpenEarth

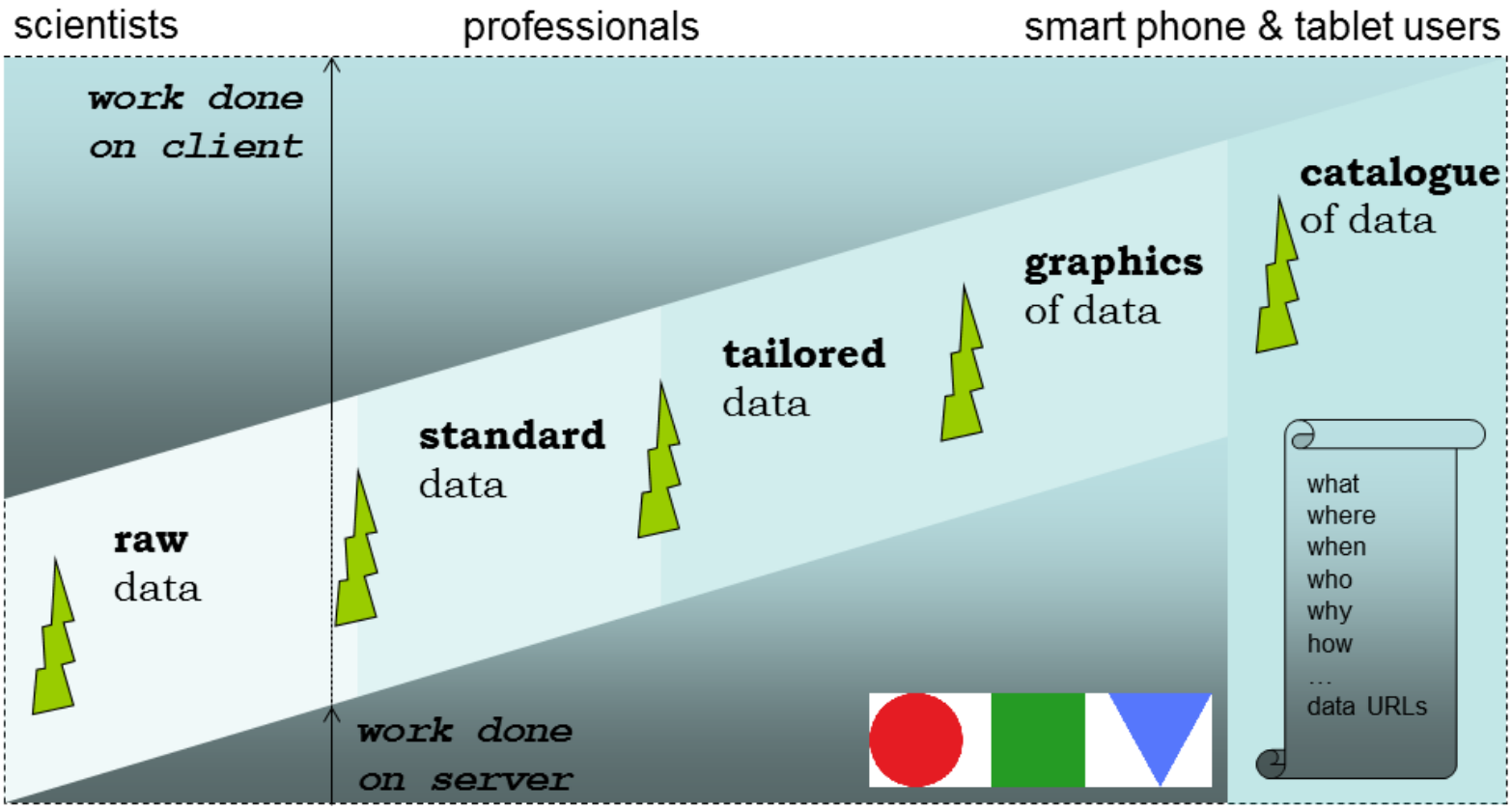
Community

- LinkedIn group (729 members)
- `http://www.openearth.eu` / wiki
- `https://github.com/openearth`
- subversion repositories (`https://svn.oss.deltares.nl/`):
  - OpenEarth Tools
  - OpenEarth Models (schematizations)
  - OpenEarth Raw Data

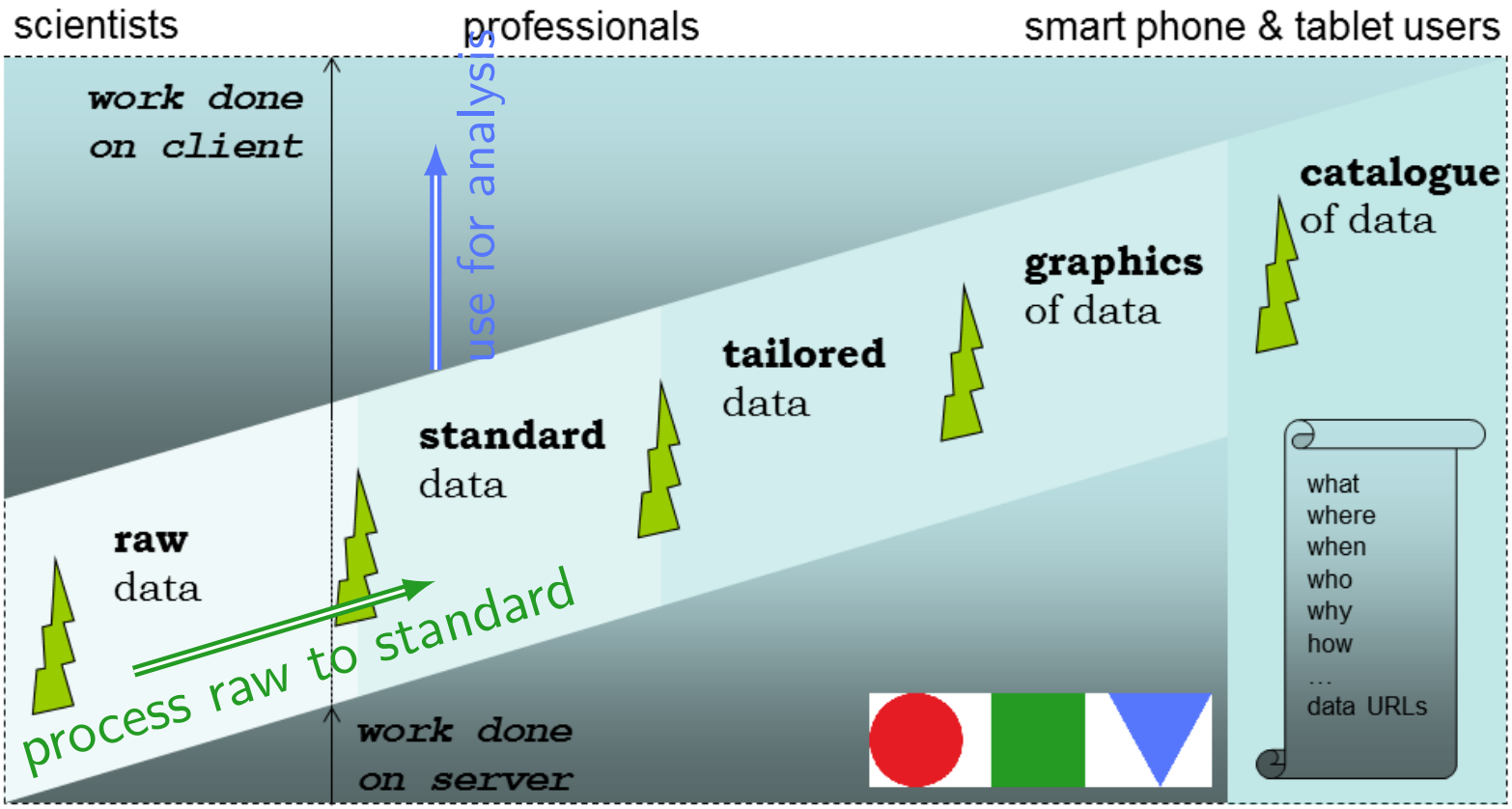
Highlights

- Open source software
- Open sharing of data, models and tools
- Reproducible workflows
- Version control (svn/git) for raw data, models and tools
- Open standards (OGC, CF, netCDF)

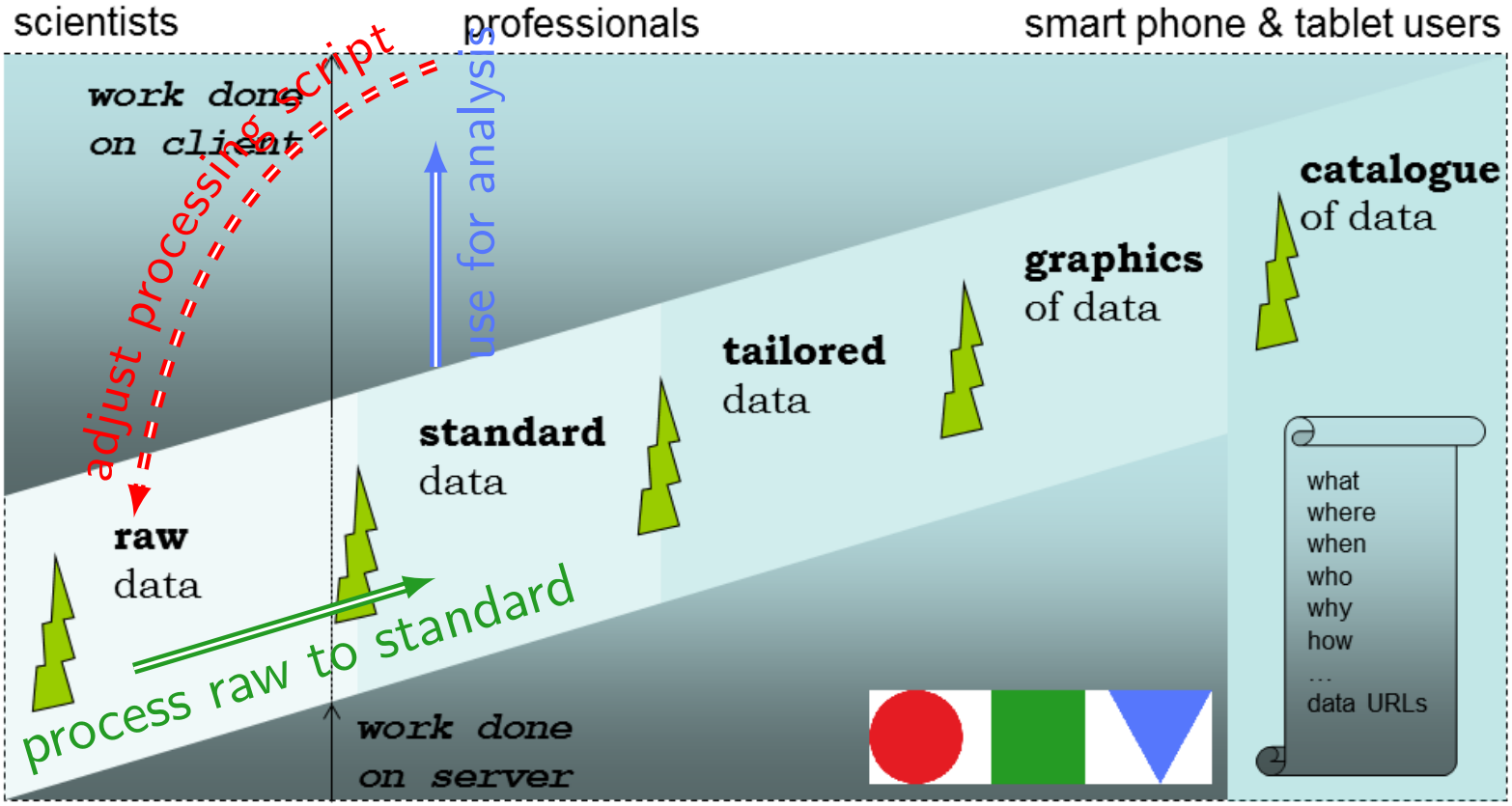
OpenEarth


scientists
professionals
smart phone & tablet users
work done on client
raw data
standard data
tailored data
graphics of data
catalogue of data
what
where
when
who
why
how
…
data URLs
work done on server

OpenEarth


scientists
professionals
smart phone & tablet users
work done on client
use for analysis
raw data
standard data
tailored data
graphics of data
catalogue of data
process raw to standard
work done on server
what
where
when
who
why
how
…
data URLs

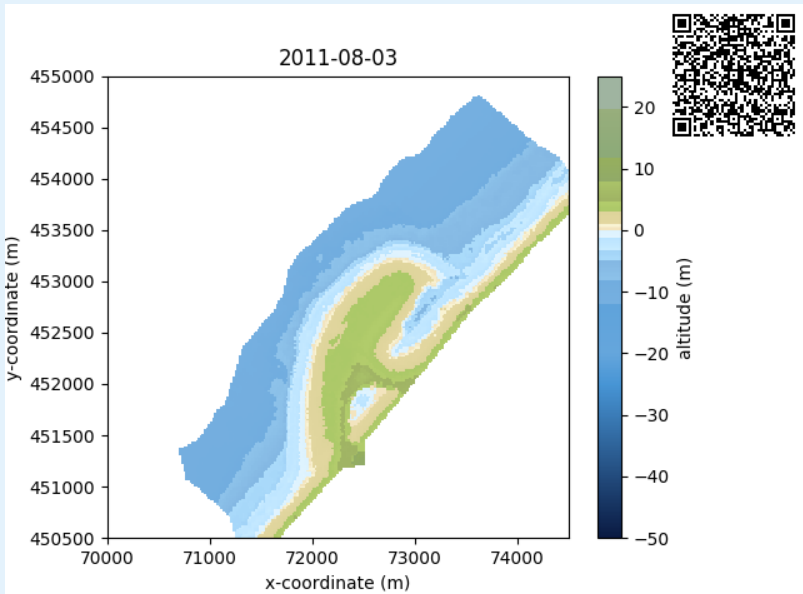
OpenEarth


scientists
professionals
smart phone & tablet users
work done on client
adjust processing script
use for analysis
raw data
standard data
tailored data
graphics of data
catalogue of data
what
where
when
who
why
how
…
data URLs
process raw to standard
work done on server

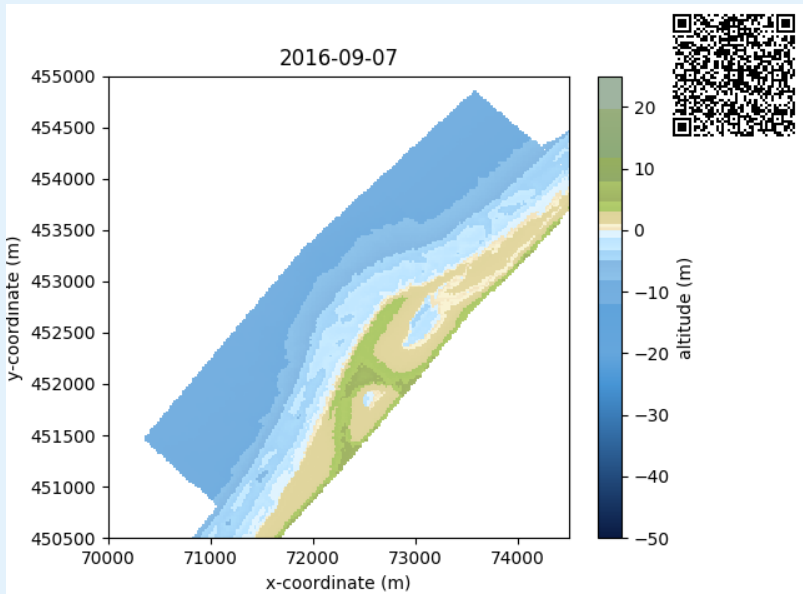
Graphics

Sand motor topography/bathymetry

2011



2016



Graphics

How created, with reference to source data

zandmotor_bathy.py Revision 14141

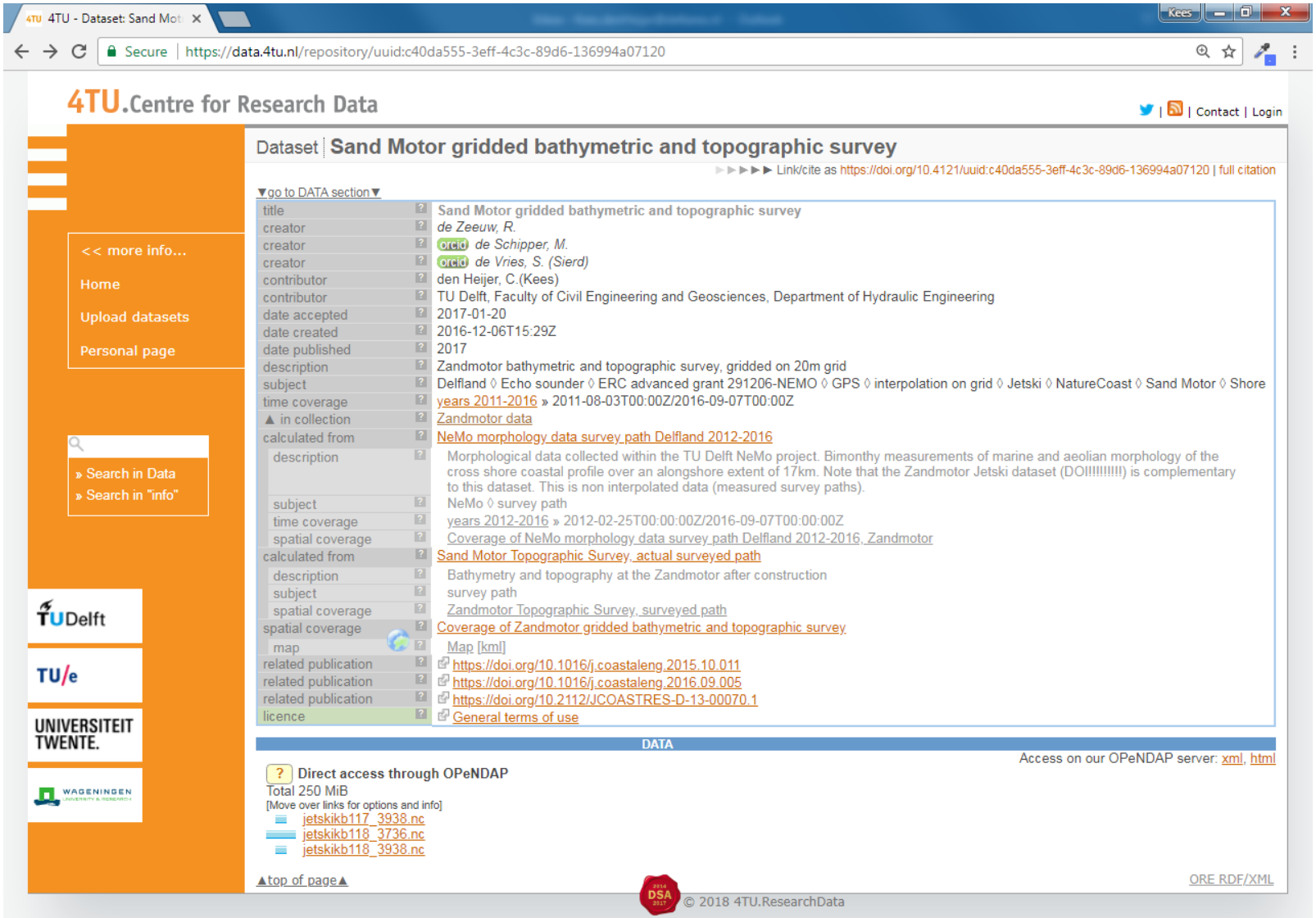
```
# -*- coding: utf-8 -*-
"""

$Id: zandmotor_bathy.py 14141 2018-02-02 14:12:01Z heijer $
$Date: 2018-02-02 15:12:01 +0100 (Fri, 02 Feb 2018) $
$Author: heijer $
$Revision: 14141 $
$HeadURL: https://svn.oss.deltares.nl/repos/openearthtools/trunk/python/OpenEarthTools/openearthtools/plot/zandmotor_bathy.py $

@author: heijer
"""

import netCDF4
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
import qrcode

from colormap_vaklodingen import vaklodingen_colormap

# https://doi.org/10.4121/uuid:c40da555-3eff-4c3c-89d6-136994a07120
nc_url = 'http://opendap.tudelft.nl/thredds/dodsC/data2/zandmotor/morphology/JETSKI/gridded/jetskikb118_3736.nc'
tidx = 0


def label(netcdf_var):
    longname = netcdf_var.long_name
    units = netcdf_var.units
    return '%s (%s)' % (longname, units)


def add_qrcode2fig(fig, size=.18):
    keywords = ['$HeadURL: https://svn.oss.deltares.nl/repos/openearthtools/trunk/python/OpenEarthTools/openearthtools/plot/zandmotor_bathy.py $', '$Revision: 14141 $', ]
    for i,keyword in enumerate(keywords):
        keywords[i] = keyword.replace('$', '').replace('Revision: ', '').replace('HeadURL: ', '').strip()
    img = qrcode.make('%s?p=%s' % (keywords[0], keywords[1]))
```

zandmotor_bathy.py Revision 14142

```
# -*- coding: utf-8 -*-
"""

$Id: zandmotor_bathy.py 14142 2018-02-02 14:13:05Z heijer $
$Date: 2018-02-02 15:13:05 +0100 (Fri, 02 Feb 2018) $
$Author: heijer $
$Revision: 14142 $
$HeadURL: https://svn.oss.deltares.nl/repos/openearthtools/trunk/python/OpenEarthTools/openearthtools/plot/zandmotor_bathy.py $

@author: heijer
"""

import netCDF4
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.gridspec as gridspec
import qrcode

from colormap_vaklodingen import vaklodingen_colormap

# https://doi.org/10.4121/uuid:c40da555-3eff-4c3c-89d6-136994a07120
nc_url = 'http://opendap.tudelft.nl/thredds/dodsC/data2/zandmotor/morphology/JETSKI/gridded/jetskikb118_3736.nc'
tidx = -1


def label(netcdf_var):
    longname = netcdf_var.long_name
    units = netcdf_var.units
    return '%s (%s)' % (longname, units)


def add_qrcode2fig(fig, size=.18):
    keywords = ['$HeadURL: https://svn.oss.deltares.nl/repos/openearthtools/trunk/python/OpenEarthTools/openearthtools/plot/zandmotor_bathy.py $', '$Revision: 14142 $', ]
    for i,keyword in enumerate(keywords):
        keywords[i] = keyword.replace('$', '').replace('Revision: ', '').replace('HeadURL: ', '').strip()
    img = qrcode.make('%s?p=%s' % (keywords[0], keywords[1]))
```

Tailored data

Published with DOI



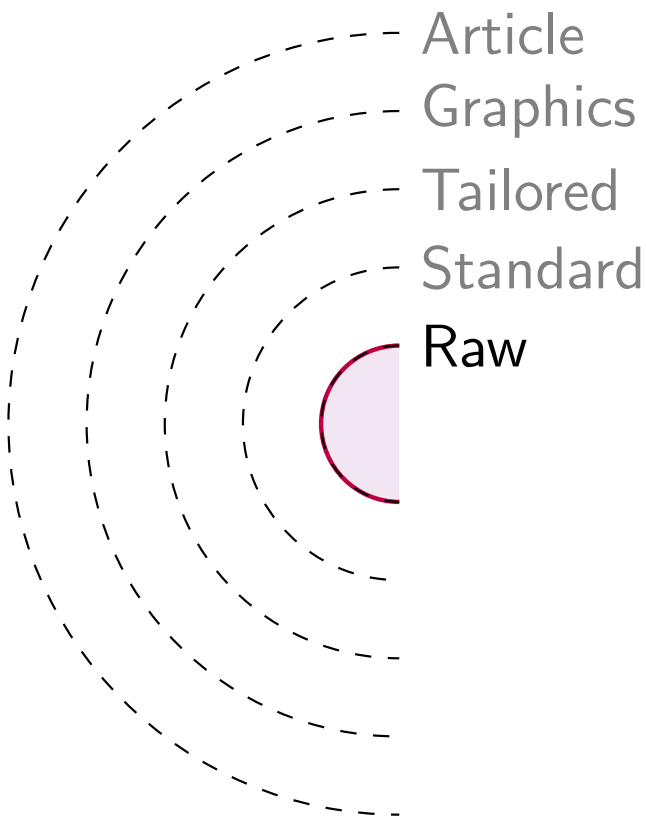
Tailored data

How created, with reference to source data

```
# -*- coding: utf-8 -*-
"""
Created on Tue Jul 29 21:38:30 2014

$Id$
$Date$
$Author$
$Revision$
$HeadURL$

@author: heijer
"""
import datetime
id='$Id$'
idsplit= id.split(' ')
frmt = '%Y-%m-%dT%H:%M:%SZ'
history='netCDF created on: '+str(datetime.datetime.utcnow().strftime(frmt))+' with script '+idsplit[1]+', script revision number is
idsplit[4] + '.'

from python.jetski_ncgen import NCgen
import os

# initialize class
nc = NCgen()

# adjust settings
nc.schema.attributes['comment'] = ''
nc.schema.attributes['history'] = history
nc.schema.attributes['license'] = 'DISCLAIMER: This data is made available in the hope that it will be useful, but WITHOUT ANY WARRA
MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.'
# Creator Search
nc.schema.attributes['creator_name'] = 'R. de Zeeuw, M. de Schipper, S. de Vries'
nc.schema.attributes['creator_url'] = 'https://www.shoremonitoring.nl and https://www.tudelft.nl'
nc.schema.attributes['creator_email'] = 'roeland@shoremonitoring.nl, m.a.deschipper@tudelft.nl, sierd.devries@tudelft.nl'
nc.schema.attributes['institution'] = 'Shore Monitoring & Research and Technische Universiteit Delft'
nc.schema.attributes['project'] = 'Zandmotor monitoring'
# Publisher Search
nc.schema.attributes['publisher_name'] = 'Kees den Heijer'
nc.schema.attributes['publisher_email'] = 'c.denheijer@tudelft.nl'
nc.schema.attributes['publisher_url'] = 'https://www.tudelft.nl'

# grid settings
nc.schema.grid.tiletype='vaklodingen'
nc.schema.grid.tileprefix = 'jetskiKB'
# path settings
nc.schema.ncpath = os.path.join(os.path.dirname(__file__), '..', '..', 'results')
if not os.path.exists(nc.schema.ncpath):
    os.mkdir(nc.schema.ncpath)
nc.raw.path = os.path.join(os.path.dirname(__file__), '..', 'raw')

# run data processing
nc.process_data()
```

What is your data management like?

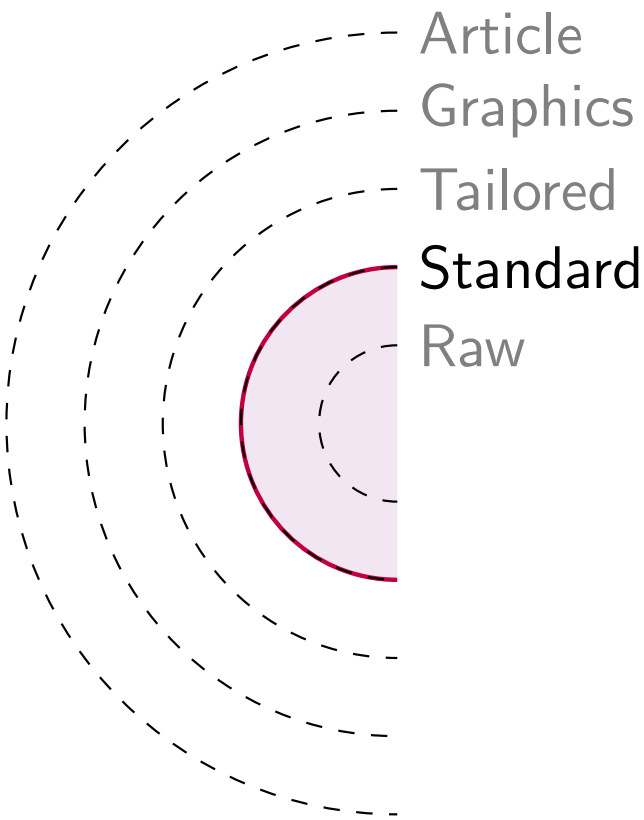
Raw data



Just as collected at the sensor.
Essential information for interpretation should be attached:

- sensor type
- location
- circumstances
- logbook

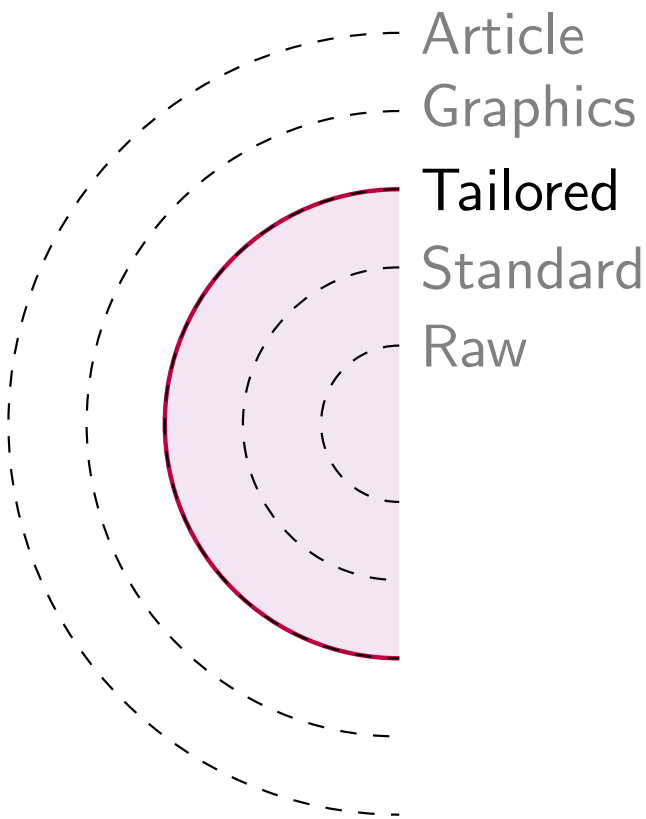
Standard data



Standardised in terms of:

- format
- units
- lineage
- metadata (vocabularies)

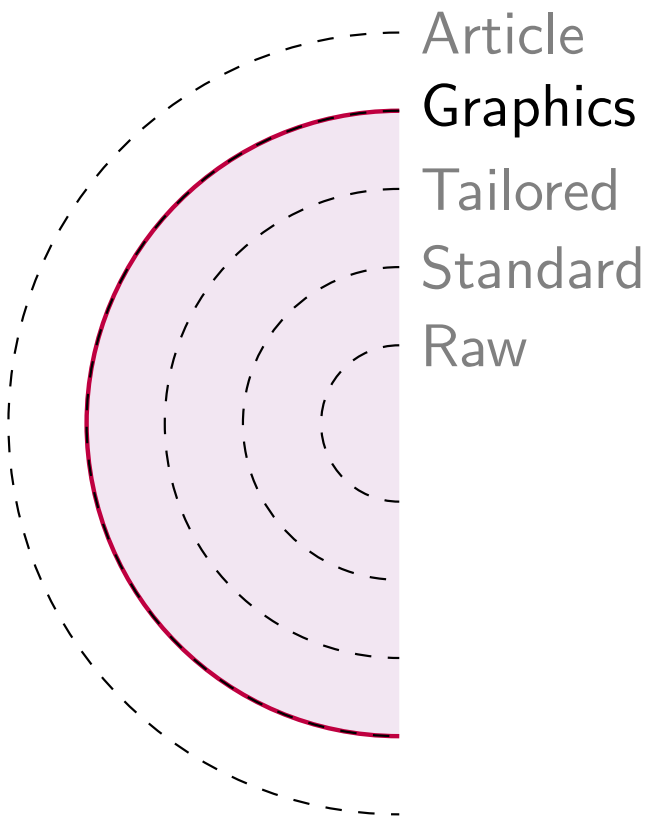
Tailored data



Tailored to particular research questions and/or (re)use

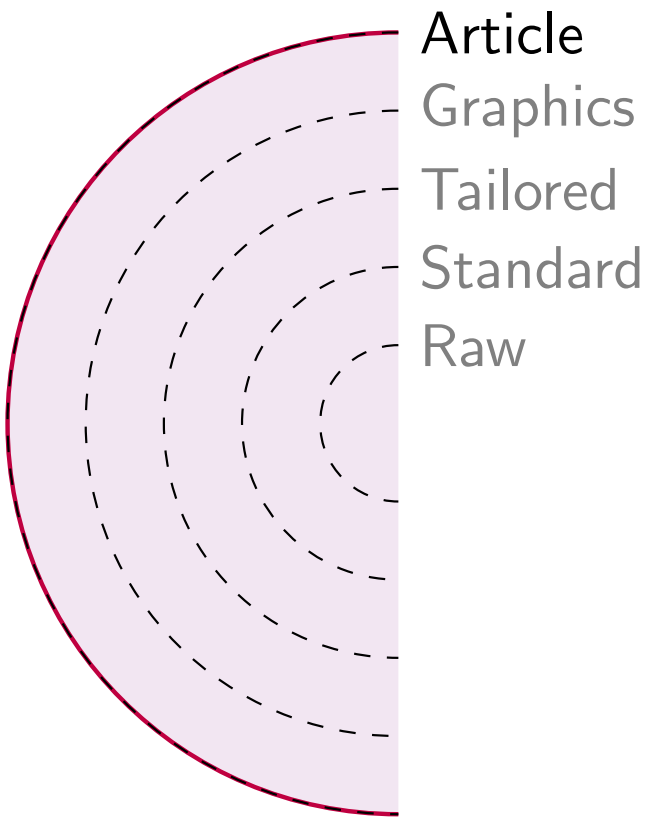
- Aggregation
- Analysis
- Combination

Graphics



Visualisation of standard and/or tailored data.
Optionally interactive.

Conclusions / Article



Wrap-up of the research, referring to all other layers and explaining how they interrelate and provide evidence for the conclusions.

Exercise

Sketch your data workflow

Include for instance:

- observation data
- processing
- interpretation

Exercise

Sketch your data workflow

Include for instance:

- observation data
- processing
- interpretation

Prioritise

Indicate key items in your workflow, based on e.g.:

- data volume
- possibility to recreate/redo
- time/resources needed to recreate/redo

Contact

Data Steward Civil Engineering and Geosciences

Questions?

name	Kees den Heijer
room	2.36 or at a CEG department
availability	Tuesday - Friday
phone	+31 15 27 85607
email	C.denHeijer@tudelft.nl / DataSteward-CiTG@tudelft.nl
ORCID	https://orcid.org/0000-0003-0314-2779
LinkedIn	https://www.linkedin.com/in/keesdenheijer/